基于用户偏好与商品属性情感匹配的图书个性化推荐研究*

侯银秀　李伟卿　王伟军　张婷婷
(华中师范大学信息管理学院　武汉　430079)
(华中师范大学青少年网络心理与行为教育部重点实验室　武汉　430079)

摘要:【目的】识别并获取细粒度的用户偏好信息, 优化图书个性化推荐的效果。【方法】使用情感分析方法对用户图书评论进行属性层文本挖掘, 通过用户本身的图书评论获取用户对图书属性的偏好; 基于每本图书的所有评论的情感计算获得其属性评分; 将用户偏好矩阵、图书属性得分矩阵进行匹配, 从而实现用户对图书属性情感偏好的个性化推荐。【结果】利用亚马逊图书评论数据作为数据来源分别对传统的协同过滤方法与本文提出的推荐方法进行实验对比。结果表明, 本文提出的方法在准确性、召回率、覆盖率上分别提高了 0.030、0.097、0.2812。【局限】未考虑时间因素对用户偏好的影响, 并且属性类型的全面程度受亚马逊图书评论数量和质量的限制。【结论】本文计算用户对图书属性的情感得分, 得到细粒度的用户偏好信息, 并通过与图书属性的得分进行匹配, 提升了图书个性化推荐的效果。

关键词: 图书个性化推荐　情感匹配　商品属性　用户偏好
分类号: G35

1　引　言

近年来, 推荐系统被证明是一种解决信息过载和长尾物品问题的有效工具, 已经与日常生活息息相关, 如在阅读新闻资讯、网上购物、聆听音乐、观看视频时, 都能看到各种各样的推荐。对于用户, 推荐系统可以帮助快速找到感兴趣的信息或物品, 从而减轻用户的认知负担, 改善用户体验, 增加用户粘性。图书个性化推荐最早是亚马逊为了提升长尾图书的用户抵达率而提出的, 据 VentureBeat 统计, 图书个性化推荐为亚马逊贡献了 35%的销售额[1]。

图书个性化推荐系统以协同过滤为主, 通过对用户–项目评分矩阵中评分数据的统计, 计算用户或项目的相似性, 将相似度高的近邻用户喜欢的图书推荐给目标用户。这种方法得到了广泛的应用, 但是仍然有一些不足: 评分数据虽然能表示用户对商品的态度, 但不能解释用户评分产生差异的详细原因; 这种方法假设给出相似评分的用户具有类似的喜好, 但近邻用户并不能完全客观、真实地反映用户自身的偏好。同时, 内容特征也会影响推荐效果, 这类推荐系统[2]从标签、评论或博客、微博等形式的用户生成内容中收集用户、商品以及用户对商品的描述信息, 并据此将用户可能喜欢的商品推荐给用户。这种方法融合了语义信息, 过滤了用户不感兴趣的图书资源, 缺点是难以区分图书内容的质量。而且, 筛选关键词和标签的做法无法识别用户的情感态度, 也无法挖掘出用户对图书细节方面的情感态度。而情感对人类认知和购买决策行为具有重要的影响和作用。Hu 等[3]研究了图书销量与用户评论数量及评论情感之间的关系, 发现在线评论数量及用户反馈中的情感体验均能影响产品

通讯作者: 王伟军, ORCID: 0000-0003-4948-0634, E-mail: wangwj@mail.ccnu.edu.cn。

*本文系国家自然科学基金项目“基于用户偏好感知的 SaaS 服务选择优化研究”(项目编号: 71271099)和国家自然科学基金项目“基于屏幕视觉热区的网络用户偏好提取及交互式个性化推荐研究”(项目编号: 71571084)的研究成果之一。

销量。Sohail 等[4]和 Diao 等[5]研究发现属性层面的情感分析方法能够提高推荐的准确性。因此，为了更深入地挖掘用户对图书属性的偏好，蕴含丰富用户意见和观点的评论内容无疑最具分析价值。用户评论中不仅包括用户提及的图书属性，还包括对属性的情感态度(关注和挑剔程度)，即通过对用户评论内容的分析，可以获取用户对图书属性的偏好，并以此为推荐依据向用户推荐更加符合其偏好的图书。刘凯等[6-7]指出准确地获取用户偏好信息是个性化推荐的基础，并提出基于商品自组织层次聚类的用户偏好表示模型。而通过对用户评论内容进行属性层情感分析，可以对用户偏好信息进行更为细粒度的解读，获取用户对商品属性的偏好。因此，在团队的前期研究基础上，为了改善推荐的效果，帮助企业更为便捷、快速地获取用户意见并提供优质的个性化推荐服务，本文筛选出影响用户偏好的图书属性，采用情感分析技术提取用户图书评论中蕴含的属性层面的观点作为推荐依据，并计算出用户偏好矩阵、图书属性得分矩阵，依据用户对图书属性的偏好和图书历史评论的情感分析结果之间的匹配程度，将最符合用户偏好的图书推荐给用户。

2 相关研究

关于图书个性化推荐的研究现状梳理主要从推荐方法和推荐依据的选择两个方面展开。

(1) 图书个性化推荐方法主要有内容推荐、协同过滤推荐和混合推荐方法等。最初的图书个性化推荐根据用户的搜索、浏览、购买、注册等行为涉及的主题信息推测用户感兴趣的书籍，推荐给用户具有相似主题或者标签的图书。但是，由于用户的能力和专业水平的差异也会造成用户产生这些行为，因此使用这种方法挖掘用户偏好会产生偏差[8]。因此要精准匹配图书资源与用户偏好，基于内容的推荐直接利用所购图书主题等信息做出推荐还有待改进。主流的协同过滤方法引入相似用户或者相似群体[9]的评分也就是观点信息，借助用户历史行为(包括浏览购买评价等)的共同性等指标，将近邻用户所购的图书列入推荐列表。这种方法通过大量评分数据计算用户或商品之间的相似性，采用聚类技术等方法改进推荐效果[10]。此外，为了克服单一推荐方式的数据不足等缺点，混合方法也应用到图书个性化推荐研究中。Najafabadi 等[11]从用户的交互记录中提取相似偏好，采用关联规则和聚类技术结合的方法从简单的数据中挖掘用户相关性，获得了较好的推荐结果。

(2) 图书个性化推荐依据的选择多以图书主题、用户评分评论、人口统计学信息、社交网络行为、时间信息为主。王伟军等[12]针对协同过滤可扩展性和数据稀疏性问题，加入了时间约束因素，以克服无法找到相似用户数据的困难。Qiu 等[13]则在评分数据基础上融入评论文本，采用评论文本的主题分布表示商品特征，挖掘评论文本在推荐系统中的价值。社交网络信息也是部分学者进行用户偏好建模考虑的因素，社交网络信息细分为推荐信任程度、社交关系等[14]，学者将这些因素融合到传统的推荐方法中，完善推荐系统依据的参数，并且提高了商品的推荐准确性。本团队在前期研究中提出，个性化推荐系统主要面向系统使用的主体——用户的实际行为与真实主观感受，以“人”的偏好为本[15]。现有图书个性化推荐的研究主要依据近邻用户的评分信息预测用户可能感兴趣的商品，未能挖掘出用户的真实主观感受。不同于评分购买记录信息，用户评论是用户自身对商品属性的主观表达。细粒度情感分析方法从短语和词级别识别情感词及其情感倾向，对提取本文需要的用户对图书属性的情感偏好信息最为合适。

此外，笔者团队在先前的研究中通过自组织聚类的方法构建即时偏好、短期偏好和长期偏好的复合模型[16]，并通过实验验证了用户偏好模型可以提高推荐效果[6]。因此，在前期研究的基础上，本文结合属性层面的情感分析技术获取更细粒度的用户偏好信息，提出一种基于用户偏好和商品属性情感得分的情感匹配推荐方法：获取图书评论数据，并使用自然语言处理工具对数据进行分析，获取影响用户对图书偏好程度的属性特征；采用情感分析技术进行观点挖掘，分别从用户自身评论和图书的历史评论中计算得到用户偏好矩阵和图书属性得分矩阵；计算用户与图书在属性层面的情感匹配程度作为推荐的依据，并采用亚马逊图书商城的实际数据验证本文提出的推荐方法的有效性。

3 商品属性与用户偏好情感匹配分析

在对用户和图书进行属性层情感匹配分析前，需要对评论数据进行清洗与处理，筛选出影响用户对图

书偏好程度的属性。常见的提取商品属性的方法有人工标注和自然语言处理两种方法。由于人工标注不适合大规模的实验文本，因此，本文采用自然语言处理工具从评论文本中抽取与图书相关的属性特征。为了获取用户偏好矩阵、图书属性得分矩阵，选用基于情感词典的情感分析方法来计算与特征词距离最近的观点词语的情感分数，通过对用户自身评论和商品历史评论的情感分析，得到用户偏好矩阵和图书属性得分矩阵，直接利用情感匹配方法计算两个矩阵的相似度，并以此为依据进行推荐。

3.1　图书评论处理

为了提取图书属性并进行情感分析，本文选用亚马逊图书商城的真实评论作为实验数据。使用斯坦福大学提供的“Stanford Large Network Dataset Collection: Amazon Reviews”数据集，由于数据集规模较大，选取评论较多的 2013 年 7 月–2014 年 7 月之间，共 36 770 条图书评论数据作为实验数据[17]。图书评论数据集的信息包括产品和用户信息、评分信息、评论文本、认为评论信息有用等内容。亚马逊图书主要使用协同过滤、热门商品和热门新品三种推荐方法。由于亚马逊图书商城要求购买之后才能够发表评论，因此获取的评论信息具有较高可靠性。经过对无效数据和重复数据的清洗和整理，共计得到有效评论文本 220 459 条，图书数量 6 506 本，用户 21 096 名。每本图书的平均评论数量约为 34 条，每个用户的平均评论数量约为 10 条。

商品属性提取方法主要有人工标注和自然语言处理工具两种。李实等[18]人工标注了数码产品和图书等商品的评论文本，根据最小最大覆盖原则建立能覆盖评论数据集的最小属性集合，但是这种方法处理数据尤其是大规模评论费时费力。He 等[19]主要采用依存句法方式，采用词性标注提取商品属性。Sohail 等[4]同样使用依存句法分析提取图书的属性并对比权重，总结为“Occurrence”、“Helpfulness”、“Material”、“Availability”、“Irrelevancy”、“Price”等 6 类属性。这种方法可以从数据中充分挖掘出用户提及的商品属性，故本文使用英文分词工具 CoreNLP① (Java 自然语言分析库)中的分词和词性标注工具处理评论文本。笔者发现，除了“mind”、“topic”等名词或动名词形式外，图书评论者还经常使用“difficult”、“useful”等形容词表示难度、实用性等方面属性。因此，本文将属性词常见词性(名词、动名词、名词短语、形容词等)抽取出来，作为图书属性词集的候选词。使用 WordNet 2.1 版本②提供的同义词集合 Synset 扩充候选词集，减少由于词性限制造成的误差；评论文本中经常使用不同词语描述相同图书属性，需要将同类图书属性加以合并以便后续情感分析。公式(1)为 WordNet 语义相似度计算公式，以此合并图书领域属性词。

$$similarity(W_1,W_2)=\frac{\sum_{i\in(1,\cdots,|SW_1|)}^{\max_{j\in(1,\cdots,|SW_2|)}}(similarity(SW1_i,SW2_j))+\sum_{i\in(1,\cdots,|SW_2|)}^{\max_{j\in(1,\cdots,|SW_1|)}}(similarity(SW2_i,SW1_j))}{|SW_1+SW_2|} \quad (1)$$

其中，$|SW_1|$为文档中的意义解释个数；$|SW_2|$为文档中的意义解释个数。两个词语之间的距离越小，语义相似度越大[20]。例如，Similarity(thesis, topic) = 0.648, Similarity (useful, effective) = 0.65，设置相似度阈值为 0.6，则“thesis”和“topic”合并为一类特征，“useful”和“effective”合并为一类，相似的描述可以有“spirit”、“soul”、“belief”、“heart”等。这些词语表达形式不同，但是描述的含义是基本相同的。根据分词工具 CoreNLP 从评论文本提取出候选词后，笔者利用语义相似度计算来确保属性候选词分类的合理性。

表 1 为最后汇总出用户在实验评论文本中所提及的 7 大类属性，分别为“内容和主题思想(mind)”、“结构和形式(structure)”、“实用性(practice)”、“趣味性(interest)”、“难度和专业性(difficulty)”、“价格(cost)”、“质量(quality)”。

①https://stanfordnlp.github.io/CoreNLP/.

②http://wordnet.princeton.edu/wordnet/download/old-versions/.

表 1　图书属性词表

编号	类型	图书属性词
1	内容和主题思想(mind)	mind, content, thesis, topic, thought, story, setting, plot, detail, spirit, soul, idea, belief, concept, ideal, sensation, heart, thinking, thinker, theory, event, deeds, reflections, feel, feeling, view, emotion, essence, mood, humanity, characters, memories, opinion
2	结构和形式(structure)	structure, framework, layout, chapter, length, clue, thread, passages, circus
3	实用性(practice)	practice, purpose, use, useful, information, device, advice, technique, effective, creative, meaningful, impact, progress, discoveries
4	趣味性(interest)	hobby, interest, interested, interesting, moved, exciting, excite, excited, delight, delightful, surprised, delighted, pleasure, joy, joyous, joys, enjoy, enjoyable, taste, enthusiast, pleasure
5	难度和专业性(difficulty)	depth, difficult, difficulty, classic, readability, specialty, profession, major
6	价格(cost)	price, cost, value
7	质量(quality)	quality, hardcover, paperback, package, paper, cover, print, printed

3.2　情感分析方法

对于图书这类商品，不仅属性词会有专有名词，评价词语也会由于领域的专业性而产生特定的词汇。提取评价词语的效果受情感词典影响较大。本文选取WordNet辞典生成的情感词典SentiWordNet[21]，该词典通过对 WordNet 中的词条进行情感分类，标注出属于positive和negative类别的每个词条的权重大小。此外，情感词典还包括象征着情感强度的程度词，例如“very”、“so much”，还有表达着情感倾向的否定词语，例如“not”、“nothing”等。因此，使用词性标注工具将图书评论中涉及到形容词、带有否定倾向的形容词组加入到情感词典中，扩大了情感词典的规模以及在图书领域的专业性。根据扩展后的词典，将图书领域7类属性的观点态度划分为5类情感等级，分别是Very Negative、Negative、Neutral、Positive、Very Positive。这 5 类情感词语的情感等级和对应的情感极性如表2所示，例如等级2代表中立性的情感，极性为0。

表 2　情感词汇等级分类

词语	等级	极性
Very Negative	0	–1
Negative	1	–1
Neutral	2	0
Positive	3	1
Very Positive	4	1

3.3　情感匹配推荐

基于用户偏好与图书属性情感匹配的图书个性化推荐方法的主要思想是：如果用户对图书属性的评分较低，证明用户对这一属性较为挑剔，且绝对值越大，用户的挑剔程度越高。如果用户对图书某一属性的评分较高，表示用户对这一属性有容忍度，且绝对值越大，容忍度越高。推荐流程如图1所示。

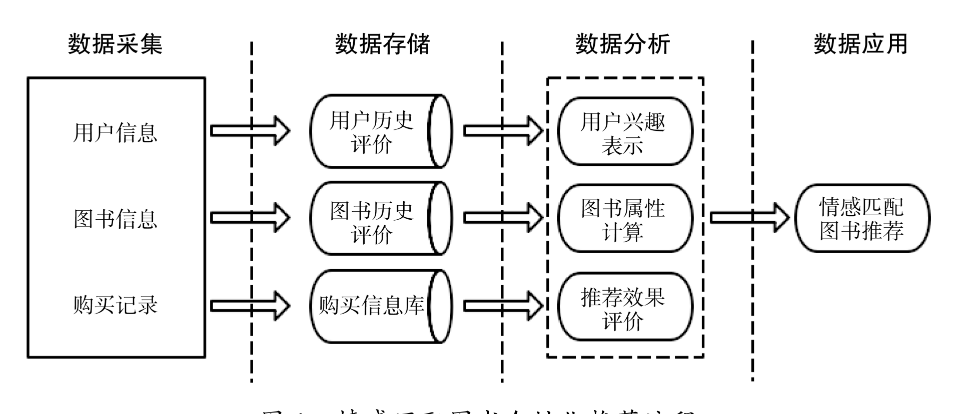


图 1　情感匹配图书个性化推荐流程

(1) 数据采集和存储阶段，收集用户和图书的相关信息。用户和图书信息包括用户和图书的基本信息及评论信息。将用户和图书的基本信息和历史评价存储到数据库中；而购买记录存储的购买信息数据库则作为后续推荐效果的评价依据。

(2) 数据分析阶段，需要对采集到的用户和图书评论信息进行细粒度情感分析。借助情感分析词典和自然语言处理工具对评论信息进行处理，获取用户偏好和图书属性的情感得分。

(3) 数据应用阶段，将用户偏好矩阵和图书属性得分矩阵进行匹配计算。将用户 u 对商品属性 t 的偏好记为向量 $U_{ui}=(U_{u1},U_{u2},U_{u3},U_{u4},\cdots)$，$U_{ui}$ 表示用户 u 对第 i 个特征的偏好程度。商品 p 的属性 t 的情感得分为 $P_{pi}=(P_{p1},P_{p2},P_{p3},P_{p4},\cdots)$，$P_{pi}$ 表示商品 p 的第 i 个特征的情感得分。本文采用余弦相似度[22]衡量用户和图书属性层的情感匹配程度，余弦相似度计算可以从方向上区分用户和图书之间的差异，通过减去均值修正用户和图书之间的度量标准不统一的问题。在此采用 $sim(X,Y)$表示用户和商品之间的情感匹配值，如公式(2)所示。

$$sim(X,Y)=\cos\theta=\frac{\overline{U_{ul}}\cdot\overline{P_{pl}}}{\|U_{ui}\|\cdot\|P_{pi}\|} \tag{2}$$

根据公式(2)的计算结果，选取与用户偏好最接近的 k 本图书，并将其推荐给用户。

4 实验与讨论

4.1 用户偏好计算

为了识别用户对图书属性的偏好，对用户自身评论数据中所有用户情感值进行汇总统计。由于本文分析的是不同用户各自的兴趣观点，因此认为所有观点持有者的评论权重是相等的。针对识别出来的图书属性词，采用 SentiByTerm 算法计算最近情感词语的得分。本文根据第 3 节提出的情感分类方法进行用户属性情感倾向值的求解。用户偏好的表示转化为情感汇总计算任务，也就是计算观点持有者某个属性类对应的情感平均值，计算方法如公式(3)所示。

$$\overline{U_{t_i}}=\frac{\sum_{j=1}^{n_{t_i}}U_{t_{i,j}}}{n_{t_i}} \tag{3}$$

其中，t_i 为属性类，n_{ti} 为属性类 t_i 在某用户历史评论中的总次数，$U_{t_{i,j}}$ 为评论中第 j 次出现的属性类 t_i 对应的情感倾向值，$\overline{U_{t_i}}$ 为某个用户所有评论中的属性类所对应的平均情感倾向值。用户为 u，属性类用 t 表示，用户偏好矩阵如图 2 所示。

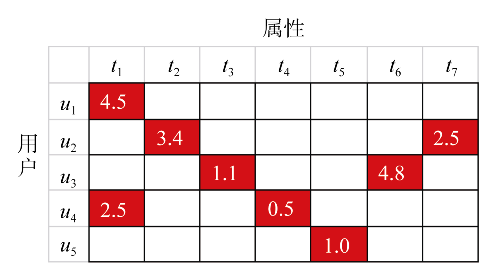


图 2 用户偏好矩阵

用户偏好矩阵是指用户对图书属性的偏好程度。为了便于采用上述情感匹配公式，将计算所得情感得分减 2，即可得到代表正负极性的情感得分。如果值大于 0，那么用户对该产品特征持有肯定态度，小于 0 则证明该产品得到的评价是否定的，值为 0 则表示持中立态度。经过计算得到结果如表 3 所示。

表 3 用户偏好表示(部分用户偏好)

用户	t_1 (内容)	t_2 (结构)	t_3 (实用性)	t_4 (趣味性)	t_5 (专业性)	t_6 (价格)	t_7 (质量)
u_1	1.56	0.00	0.83	2.63	1.18	0.00	1.00
u_2	–1.00	0.53	0.76	0.00	1.42	1.56	–1.83
u_3	1.00	1.89	–2.00	0.00	0.00	1.56	0.34
u_4	0.00	2.00	–1.82	0.34	1.88	1.57	–2.00
u_5	0.80	–2.00	1.55	0.59	0.35	2.00	–2.00

4.2 图书属性情感分析

为了评估推荐效果，对图书属性情感得分加以计算。商品属性得分计算如公式(4)所示。

$$\overline{P_{t_i}}=\frac{\sum_{j=1}^{n_{t_i}}P_{t_{i,j}}}{n_{t_i}} \tag{4}$$

其中，t_i 为属性类，n_{t_i} 为属性类出现的总次数，$P_{t_{i,j}}$ 为商品历史评论中第 j 出现的属性类 t_i 对应的情感倾向值，$\overline{P_{t_i}}$ 为某个商品所有评论中的属性类所对应的平均情感倾向值。商品用 p 表示，属性类用 t 表示，图书属性得分矩阵如图 3 所示。

图书属性矩阵就是计算所有参与评论用户对每个图书属性的情感倾向。将产品的情感得分减 2，即可得到能反映图书属性情感得分的矩阵。结果如表 4 所示。

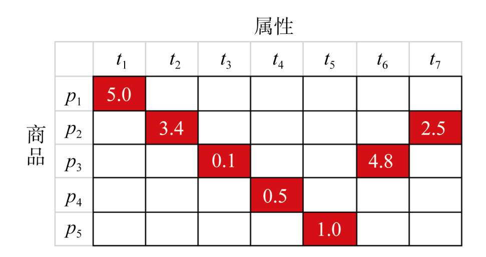


图 3　图书属性得分矩阵

表 4　图书属性情感分析(部分图书得分)

图书	t_1 (内容)	t_2 (结构)	t_3 (实用性)	t_4 (趣味性)	t_5 (专业性)	t_6 (价格)	t_7 (质量)
p_1	–0.56	0	–0.29	1	–1	0	1
p_2	0.14	0	–0.57	0.44	0.33	1	0
p_3	–0.34	0	–0.67	0	0	1	–0.4
p_4	–0.21	–0.08	–0.68	–2	–1	0	–0.67
p_5	0.2	0	–0.33	0	0.5	0	0
p_6	0	0	–0.17	0	0	0	–0.67
p_7	–0.49	0	0.95	0	–1	1	0
p_8	–0.4	0	–1.33	1	–1	0	–1
p_9	–0.13	0	–0.71	0.07	1	0	0

4.3　推荐结果分析与讨论

根据用户自身评论和商品历史评论分别计算用户偏好矩阵和图书属性得分矩阵。用户偏好矩阵大小为 m×t，m 代表用户数量，t_i 代表商品属性类别，$\overline{U_{t_i}}$ 表示用户 u 对属性 t_i 的平均情感倾向值，如果用户没有对该类属性进行评价，则分数取 0。图书属性得分矩阵大小为 n×t, n 代表商品数量，t_i 代表商品属性类别，$\overline{P_{t_i}}$ 代表商品 p 在属性类 t_i 上的情感评分。

利用用户偏好矩阵和图书属性情感得分矩阵进行相似度计算，计算出与待推荐用户相似度 Top10 的商品，得到推荐结果。实验效果的测评采用准确率、召回率与覆盖率三个指标[23]。本实验将数据集划分为训练集和测试集两个部分，训练集用来产生推荐结果，测试集对结果进行比较与评估。在根据用户偏好与商品属性的情感匹配进行推荐实验后，另外采用传统的基于项目的推荐(Item-base)和基于用户的推荐(User-base)进行比较。为了避免推荐系统中数据稀疏问题，影响实验比较，本文选用评论总数大于 30 的用户数据，共有 163 位用户和 4 287 本图书，26 160 条评论数据进行实验。表 5 为三种方法推荐效果的比较。

表 5　Item-base、User-base 与基于用户偏好与商品属性的情感分析推荐算法结果对比

推荐算法	正确推荐数量	推荐数量	准确率	召回率	覆盖率
Item-base	244	1 630	0.1497	0.1124	0.3418
User-base	249	1 630	0.1528	0.1285	0.3051
本文方法	298	1 630	0.1828	0.1382	0.5863

图书个性化推荐采用基于用户偏好与商品属性的情感分析的推荐算法相对于传统的 Item-base 和 User-base 将推荐的准确率提高了 0.030，召回率提高了 0.097，覆盖提高了 0.2812。在基于用户偏好与商品属性的推荐算法中，不需要寻找相似用户或类似商品，而是将用户的情感倾向和商品的情感得分直接匹配，筛选出满足用户偏好的商品。由于本文在基于项目的协同过滤方法基础上，通过挖掘用户对图书属性的偏好，使得推荐结果更加符合用户需求。对于评论较少的用户而言，采用的是亚马逊的“认为此评论有用”原则，将用户赞同过的评论作为用户的观点和偏好，减少冷启动问题带来的偏差。从实验结果来看，本文提出的基于用户偏好和商品属性的推荐算法提高了图书个性化推荐效果。

5 结 语

传统的图书个性化推荐主要依据近邻用户的评分信息做出推荐，无法完全客观、真实地表示用户自身的偏好以及详细的原因。虽然有学者结合图书主题、作者、体裁等属性研究用户偏好，但是没有深入挖掘用户的意见和情感态度。本文采用情感分析技术从商品属性层面获取用户偏好和图书在不同属性方面的情感得分，对用户和图书进行属性层情感匹配并根据匹配结果做出推荐。

本文对用户自身评论和商品历史评论进行属性层面的情感匹配分析，为图书个性化推荐提供了一种新的方法。在梳理现有图书个性化推荐方法和推荐依据参数的基础上，本文总结了现有图书个性化推荐研究存在的不足之处；获取用户对图书属性特征的偏好，对用户偏好进行更细粒度的刻画；为降低依据近邻用户偏好产生推荐的误差，使用情感分析方法构建用户偏好矩阵和图书属性得分矩阵，提出了依据用户偏好和商品属性情感得分直接匹配的推荐方法，提升了推荐效果，并为网络书店图书个性化推荐提供了新的思路。

本研究也存在一些不足：本文将图书属性汇总归纳为 7 类，但商品属性尚不够完善，需要更加完善地挖掘影响用户偏好的图书属性；用户偏好会随着时间的推移而发生变化，没有考虑时间因素对用户偏好的影响；受评论数据量和质的影响较大，建议与协同过滤方法结合使用效果更佳；在计算图书属性得分矩阵时，没有考虑用户的权重。而初级用户和有经验用户发表的意见权威性有所不同，用户的意见权重因素将是下一步探索的内容。

参考文献：

[1] 刘建国，周涛，汪秉宏. 个性化推荐系统的研究进展[J]. 自然科学进展，2009, 19(1): 1-15. (Liu Jianguo, Zhou Tao, Wang Binghong. Research Progress of Personalized Recommendation System[J]. Progress in Natural Science, 2009, 19(1): 1-15.)

[2] Xu Y, Yin J. Collaborative Recommendation with User Generated Content[J]. Engineering Applications of Artificial Intelligence, 2015, 45: 281-294.

[3] Hu N, Koh N S, Reddy S K. Ratings Lead You to the Product, Reviews Help You Clinch It? The Mediating Role of Online Review Sentiments on Product Sales[J]. Decision Support Systems, 2014, 57(1): 42-53.

[4] Sohail S S, Siddiqui J, Ali R. Book Recommendation System Using Opinion Mining Technique[C]//Proceedings of the 2013 International Conference on Advances in Computing, Communications and Informatics. IEEE, 2013: 1609-1614.

[5] Diao Q, Qiu M, Wu C Y, et al. Jointly Modeling Aspects, Ratings and Sentiments for Movie Recommendation (JMARS)[C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2014: 193-202.

[6] Liu Q, Liu K, Wang Y, et al. User-Oriented Real-Time Preference Obtaining Screen Visual Hotspots Recognition Based on the Research of Eye Movement Experiments[C]// Proceedings of the 1st International Conference on Information Acquisition and Knowledge Services.2014, 39: 181-185.

[7] 刘凯. 基于屏幕视觉热区的用户偏好提取及个性化推荐[M]. 北京：科学出版社，2016. (Liu Kai. User Preference Extraction and Personalized Recommendation Based on Screen Visual Hotspot[M]. Beijing: Science Press, 2016.)

[8] 李树青，徐侠，许敏佳. 基于读者借阅二分网络的图书可推荐质量测度方法及个性化图书推荐服务[J]. 中国图书馆学报，2013, 39(3): 83-95. (Li Shuqing, Xu Xia, Xu Minjia. The Measures of Books' Recommending Quality and Personalized Book Recommendation Service Based on Bipartite Network of Readers and Books' Lending Relationship [J]. Journal of Library Science in China, 2013, 39(3): 83-95.)

[9] 林耀进，胡学钢，李慧宗. 基于用户群体影响的协同过滤推荐算法[J]. 情报学报，2013, 32(3): 299-305. (Lin Yaojin, Hu Xuegang, Li Huizong. Collaborative Filtering Recommendation Algorithm Based on User Group Influence[J]. Journal of the China Society for Scientific and Technical Information, 2013, 32(3): 299-305.)

[10] Liao C L, Lee S J. A Clustering Based Approach to Improving the Efficiency of Collaborative Filtering Recommendation [J]. Electronic Commerce Research & Applications, 2016, 18: 1-9.

[11] Najafabadi M K, Mahrin M N, Chuprat S, et al. Improving the Accuracy of Collaborative Filtering Recommendations Using Clustering and Association Rules Mining on Implicit Data[J]. Computers in Human Behavior, 2017, 67: 113-128.

[12] 王伟军，宋梅青. 一种面向用户偏好定向挖掘的协同过滤个性化推荐算法[J]. 现代图书情报技术，2014(6): 25-32.

(Wang Weijun, Song Meiqing. A Collaborative Filtering Personalized Recommendation Algorithm Through Directionally Mining User' Preferences[J]. New Technology of Library and Information Service, 2014(6): 25-32.)

[13] Qiu L, Gao S, Cheng W, et al. Aspect-based Latent Factor Model by Integrating Ratings and Reviews for Recommender System[J]. Knowledge-Based Systems, 2016, 110: 233-243.

[14] Li Y M, Wu C T, Lai C Y. A Social Recommender Mechanism for E-commerce: Combining Similarity, Trust, and Relationship[J]. Decision Support Systems, 2013, 55(3): 740-752.

[15] 刘凯, 王伟军, 黄英辉, 等. 个性化推荐系统理论探索: 从系统向用户为中心的演进[J]. 情报理论与实践, 2016, 39(3): 52-56. (Liu Kai, Wang Weijun, Huang Yinghui, et al. Exploration on Personalized Recommendation System Theory: Evolution from System-centric to User-centric[J]. Information Studies: Theory & Application, 2016, 39(3): 52-56.)

[16] 丁雪. 认知视角下网络用户偏好复合模型的构建与验证[D]. 武汉: 华中师范大学, 2016. (Ding Xue. A Composite User Preference Model: An Experimental Model from a Cognitive Perspective[D]. Wuhan: Central China Normal University, 2016.)

[17] Mcauley J, Leskovec J. Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text[C]// Proceedings of the 7th ACM Conference on Recommender Systems. 2013: 165-172.

[18] 李实, 叶强, 李一军, 等. 中文网络客户评论的产品特征挖掘方法研究[J]. 管理科学学报, 2009, 12(2): 142-152. (Li Shi, Ye Qiang, Li Yijun, et al. Mining Features of Products from Chinese Customer Online Reviews[J]. Journal of Management Sciences in China, 2009, 12(2): 142-151.)

[19] He T, Hao R, Qi H, et al. Mining Feature-Opinion from Reviews Based on Dependency Parsing [J]. International Journal of Software Engineering & Knowledge Engineering, 2016, 26(10): 1581-1591.

[20] Lopez-Arevalo I, Sosa-Sosa V J, Rojas-Lopez F, et al. Improving Selection of Synsets from WordNet for Domain-specific Word Sense Disambiguation[J]. Computer Speech & Language, 2017, 41(10): 128-145.

[21] Hung C, Chen S J. Word Sense Disambiguation Based Sentiment Lexicons for Sentiment Classification [J]. Knowledge-Based Systems, 2016, 110(5): 224-232.

[22] 汤颖, 钟南江, 范菁. 一种结合用户评分信息的改进好友推荐算法[J]. 计算机科学, 2016, 43(9): 111-115. (Tang Ying, Zhong Nanjiang, Fan Jing. An Improved Friend Recommendation Algorithm Combined with User Rating Information[J]. Computer Science, 2016, 43(9): 111-115.)

[23] 项亮. 推荐系统实践[M]. 北京: 人民邮电出版社, 2012. (Xiang Liang. Recommendation System Practice [M]. Beijing: Posts & Telecom Press, 2012.)

作者贡献声明:

侯银秀: 构建推荐模型, 处理数据, 撰写论文;
李伟卿: 完成情感计算和推荐实验;
王伟军: 提出研究命题及研究思路, 修改论文;
张婷婷: 结论概括, 论文修改。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据见期刊网络版 http://www.infotech.ac.cn。
[1] 侯银秀. 图书情感得分.xlsx. 图书评论情感得分.
[2] 侯银秀. 用户情感得分.xlsx. 用户评论情感得分.
[3] 侯银秀. 推荐结果.txt. 推荐结果.

收稿日期: 2017-05-22
收修改稿日期: 2017-07-11

Personalized Book Recommendation Based on User Preferences and Commodity Features

Hou Yinxiu　Li Weiqing　Wang Weijun　Zhang Tingting
(School of Information Management, Central China Normal University, Wuhan 430079, China)
(Key Laboratory of Adolescent Cyber Psychology and Behavior, Ministry of Education,
Central China Normal University, Wuhan 430079, China)

Abstract: [**Objective**] This paper identifies the fine-grained preferences of online bookstore users, aiming to optimize the personalized book recommendation service. [**Methods**] First, we conducted sentiment analysis of the book features through readers' comments, which indicated their preferences. Then, we calculated the books' sentiment scores based on the readers' comments. Finally, the user preferences matrix and the sentiment scores matrix were matched to personalize the book recommendation. [**Results**] We retrieved the needed data from Amazon's book comments, and then conducted an experiment to compare the results of our new method with those of the traditional collaborative filtering methods. We found that the proposed method improved the precision, recall and coverage by 0.030, 0.097, 0.2812. [**Limitations**] We did not consider the impacts of time on user's preferences, and the feature types might not be comprehensive due to the limited number and quality of Amazon's book comments. [**Conclusions**] The proposed method improves the performance of personalized book recommendation service.

Keywords: Personalized Book Recommendation　Sentiment Matching　Commodity Feature　User Preference

瑞士国家科学基金会要求申请项目必须提供数据管理计划

瑞士国家科学基金会新增一项项目经费申请要求: 从 2017 年 10 月开始, 研究人员必须提交一份数据管理计划, 这份数据管理计划将作为其完整研究计划的一部分。

数据管理计划是迈向开放科研数据的重要一步。世界范围内, 已经有一些倡议力图使科学数据, 尤其是科研数据开放可用。瑞士国家科学基金会认同这些倡议的基本原则。

数据管理计划将鼓励研究人员在项目开始前便考虑数据的生命周期。如果没有限制数据公开的相关法律、道德或版权, 他们必须在非商用的数字知识库中存储数据及元数据, 存储的格式是任何人都能找到、获取和再利用的。瑞士国家科学基金会将支付数据上传的费用。

目前, 瑞士国家科学基金会正在敲定数据管理计划的指导方针。指导方针应保证每个科学学科都能够达到数据管理计划的要求。一旦方针出台, 瑞士国家科学基金会网站将会传达和发布。已在 2017 年 4 月提交审批的项目将不必提供数据管理计划。

开放数据能极大地提升科学研究的影响力、透明度和再现性。

(编译自: http://www.snsf.ch/en/researchinFocus/newsroom/Pages/new-170306-towardsopen-research-data.asp)

(本刊讯)